

Identifikation multivariater Ausreißer bei Intervalldaten

Intervalldaten kommen in unterschiedlichsten Szenarien vor, z.B. bei der Aggregation

von großen Datensätzen, bei denen die einzelnen gemessenen Variablen als intervallwertige Information bei Aggregation resultieren. Jede Variable der multivariaten Intervalldaten kann in Form von Mittelpunkt und Spannweite repräsentiert werden. Es wurden parametrische Modelle vorgeschlagen, die gemeinsame multivariate Normalverteilung von Mittelpunkt und log-Spannweite der

multivariaten Intervalldaten annehmen. Unterschiedliche Annahmen über die gemeinsame

Kovarianzmatrix machen es möglich, die Beziehungen zwischen Mittelpunkt und log-Spannweite einzuschränken. In jedem Fall kann das gemeinsame Mittel und die gemeinsame Kovarianzmatrix über die Maximum-Likelihood Methodik geschätzt werden. Für eine robuste Parameterschätzung wird hier der

Trimmed Maximum-Likelihood Schätzer genommen. Multivariate Ausreißer werden schließlich mit der Mahalanobis-Distanz basierend auf den robusten Schätzern identifiziert.